Variabili casuali e distribuzioni di probabilità

Simulazione – Lezione n. 7

Corso di Laurea in Informatica Applicata
Università di Pisa, sede di La Spezia

Contenuti

- Aspetti probabilistici della simulazione
- Variabili casuali
- Definizione matematica
- Funzioni di distribuzione e di densità
- Media, mediana, moda, varianza

Determinismo della simulazione

- Esperimenti di simulazione
  - Esecuzioni di un simulatore
  - Un programma deterministico
  - Dati in ingresso e dati in uscita
- Dati che entrano
  - Parametri del sistema
  - Attributi delle entità e delle loro istanze
- Dati che escono
  - Oggetto dello studio, non noti né deducibili
  - Ma deterministici rispetto ai dati in ingresso
- I risultati di una simulazione sono riproducibili

Dati da modellare

- Dati in ingresso
  - Parametri: scelti, normalmente fissi durante la simulazione
  - Attributi: valori diversi necessari durante la simulazione
- Diversi come? Come nella realtà
  - Spesso casuali, secondo regole probabilistiche (distribuzioni)
  - Controllate: da identificare e da riprodurre
- Esempi:
  - Il valore che definisce la politica dell'impiegato
  - I tempi di arrivo e di servizio di ogni cliente
  - Il tempo di attesa e il giudizio di posizione dei clienti
  - I tempi per le pause dell'impiegato

Dati da studiare

- Dati in uscita
  - Misure del comportamento del sistema
  - Dipendenti dai parametri ...
  - ... spesso dalle condizioni iniziali del sistema ...
  - ... e dai valori assunti dagli attributi durante la simulazione
- Studio probabilistico
  - "Probabilità" in ingresso implicano "probabilità" in uscita
  - Basato su dati di più simulazioni da studiare aggregati
  - Una o poche esecuzioni sono casi particolari
  - Il tempo di produzione dei dati può essere importante
- Stessi strumenti, fini diversi

Problemi e attività

- Nel processo di simulazione
- Identificare gli attributi interessanti
  - Specificare i valori, come intervalli e come regole
  - Stime o studio di campioni di dati con strumenti probabilistici
- Riprodurli durante la simulazione
  - Generandone i valori secondo le regole specificate
  - In modo deterministicamente riproducibile
- Studiare i risultati
  - Collezionarli nella necessaria quantità
  - Studiarne le proprietà con strumenti probabilistici

Variabili casuali

- Nome controverso
- Variabile, in matematica
  - Un simbolo per un valore che può variare
  - In genere il valore è oggetto di scelta, es. x in $f(x)$
- Variabile, in informatica
  - Strumenti sintattici e semantici per rappresentare valori
  - I valori si assegnano e si leggono, es. `int x, y; ...; x = y;`
- Variabile casuale
  - Un simbolo (sintattico) il cui valore non si sceglie (assegna), ma risponde a una legge di probabilità (una distribuzione)

Variabili casuali nella simulazione

- Dati di ingresso, da modellare e riprodurre
  - Identificazione degli attributi interessanti
  - Studio dei campioni di dati per identificare le distribuzioni
  - Specifica come variabili casuali con distribuzioni definite
- Realizzazione del simulatore
  - Strumenti di supporto alle variabili casuali
  - Definizione e generazione dei valori durante la simulazione
- Dati di uscita, da studiare
  - Minimo, massimo, media, varianza, ...
  - Eventuali dipendenze dai dati iniziali
  - Valutare l'affidabilità dei risultati

Spazio di probabilità

- Una tripla (Ω, $\mathscr{F}$, P)
- Ω spazio campione
  - I possibili esiti di un esperimento
  - Es. i risultati del lancio di una moneta: $\Omega = \{T, C\}$
- $\mathscr{F}$ spazio degli eventi
  - Famiglia di sottoinsiemi di Ω
  - Es.: $\mathscr{F} = 2^\Omega = \{ \varnothing, \{T\}, \{C\}, \{T, C\}\}$
- $P : \mathscr{F} \rightarrow [0, 1]$, funzione di probabilità
  - Probabilità che gli eventi in $\mathscr{F}$ si verifichino
  - Es.: $P(\varnothing) = 0$, $P(\{T\}) = 0.5$, $P(\{C\}) = 0.5$, $P(\{T, C\}) = 1$

Spazio di probabilità, proprietà

- Per definizione
  - $\Omega \in \mathscr{F}$
  - $A \in \mathscr{F} \Rightarrow \Omega \setminus A \in \mathscr{F}$, $A, B \in \mathscr{F} \Rightarrow A \cup B \in \mathscr{F}$
  - $P(A) \geq 0$, $\forall A \in \mathscr{F}$, $P(\Omega) = 1$
  - $(A, B \in \mathscr{F}) \wedge (A \cap B = \varnothing) \Rightarrow P(A \cup B) = P(A) + P(B)$
- Derivabili, $\forall$ A, B $\in \mathscr{F}$
  - $A \cap B \in \mathscr{F}$, $B \setminus A \in \mathscr{F}$
  - $P(\varnothing) = 0$
  - $P(\Omega \setminus A) = 1 - P(A)$
  - $A \subseteq B \Rightarrow P(B \setminus A) = P(B) - P(A)$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Variabile casuale

- Dato uno spazio di probabilità (Ω, $\mathscr{F}$, P)
- Una variabile casuale reale X è una funzione
  - $X : \Omega \rightarrow \mathbb{R}$ tale che
    $\forall r \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \leq r\} \in \mathscr{F}$
- Funzione di distribuzione
  - $F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(X \leq x)$
- In pratica
  - La funzione di distribuzione caratterizza sempre una VC
  - Si "saltano" Ω e $\mathscr{F}$
  - Definendo direttamente la probabilità dei valori di X

Esempio, la VC come mappa

- Operazioni allo sportello del nostro impiegato
  - Richiesta di informazioni, prenotazione, acquisto di biglietto
  - Spazio campione: $\Omega = \{inf, pre, acq\}$
  - Spazio degli eventi: $\mathscr{F} = 2^{\Omega}$
- Definiamo la funzione: $X : \Omega \rightarrow \mathbb{R}$
  - $X(inf) = 0$, $X(pre) = 1$, $X(acq) = 2$
- Verifichiamo che $\forall r \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \leq r\} \in \mathscr{F}$
  - $r < 0$ $\quad \{\omega \in \Omega : X(\omega) \leq r\} = \varnothing$
  - $0 \leq r < 1$ $\quad \{\omega \in \Omega : X(\omega) \leq r\} = \{inf\}$
  - $1 \leq r < 2$ $\quad \{\omega \in \Omega : X(\omega) \leq r\} = \{inf, pre\}$
  - $2 \leq r$ $\quad \{\omega \in \Omega : X(\omega) \leq r\} = \{inf, pre, acq\} = \Omega$

Esempio, distribuzione della VC

- Da un campione "rappresentativo" di operazioni
  - *inf*, *pre*, *pre*, *inf*, *pre*, *acq*, *inf*, *acq*, *acq*, *inf*
  - Operazioni indipendenti, 4 *inf*, 3 *pre*, 3 *acq*
- Funzione di distribuzione, definizione
  - $F_X(x) = P(\{ \omega \in \Omega : X(\omega) \leq x \}) = P(X \leq x)$
- Funzione di distribuzione (empirica)
  - $x < 0$, $F_X(x) = P(\varnothing) = 0.0$
  - $0 \leq x < 1$, $F_X(x) = P(\{inf\}) = 0.4$
  - $1 \leq x < 2$, $F_X(x) = P(\{inf, pre\}) = 0.7$
  - $2 \leq x$, $F_X(x) = P(\{inf, pre, acq\}) = 1.0$

Variabili e distribuzioni discrete

- Variabile casuale discreta
  - $X : \Omega \rightarrow x_1, x_2, \ldots, x_k, \ldots$
  - Valori in un insieme finito o al più numerabile
  - È possibile parlare di probabilità di un valore
- Funzione di densità discreta
  - $f_X(x) = P(X = x)$ se $x = x_i$, per $i = 1, 2, \ldots k, \ldots$
  - $f_X(x) = 0$ altrimenti
- Funzione di distribuzione discreta
  - $F_X(x) = P(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$

Variabili e distribuzioni continue

- Variabile casuale continua
  - $X : \Omega \rightarrow \mathbb{R}$
  - Valori in un insieme infinito non numerabile
  - È possibile parlare solo di probabilità di valori in intervalli
- Funzione di distribuzione continua
  - $F_X(x) = P(X \leq x)$
- Relazioni fra funzioni di distribuzione e di densità

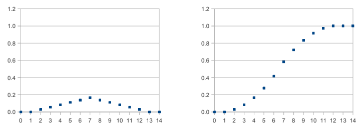
$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du \qquad f_X(x) = \frac{dF_X(x)}{dx}$$

Terminologia

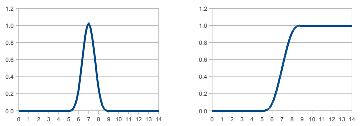
- Distribuzione
  - Termine generico per la legge seguita dai valori di una VC
  - Spesso usato per indicare la VC stessa
- Funzioni associate
  - $F_X(x)$: funzione di distribuzione, più precisamente: *cumulative probability distribution function*, per VC continue e discrete
  - $f_X(x)$: funzione di densità, più precisamente: *probability mass function*, per VC discrete, *probability density function*, per VC continue

Grafici di $f_X(x)$ e $F_X(x)$

- VC discrete



- VC continue



Informazioni interessanti su una VC

- Intervallo di definizione
- Media (o valore atteso)
  - La media dei valori della VC pesata sulla loro probabilità
- Mediana
  - Il valore che separa le metà alta e bassa dei valori della VC
- Moda
  - Il valore più probabile della VC
- Varianza (σ^2) e deviazione standard (σ)
  - Misure della dispersione fra i valori della VC

Media

- VC discrete
  $E[X] = \mu_X = \sum_i x_i f_X(x_i)$
- VC continue
  $E[X] = \mu_X = \int_{-\infty}^{+\infty} x f_X(x) dx$
- Proprietà
  $E[c] = c$ $\quad E[cX] = cE[X]$
  $E[X+c] = E[X]+c$ $\quad E[X+Y] = E[X]+E[Y]$

Varianza

- VC discrete
  $\sigma_X^2 = Var[X] = \sum_i (x_i - \mu_X)^2 f_X(x_i)$
- VC continue
  $\sigma_X^2 = Var[X] = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f_X(x) dx$
- Definizione generale
  $\sigma_X^2 = E[(X-\mu_X)^2] = E[X^2 - 2\mu_X X + \mu_X^2]$
  $= E[X^2] - 2\mu_X E[X] + \mu_X^2 = E[X^2] - \mu_X^2$
  $= E[X^2] - E[X]^2$

Perché importanti

- Intuitivamente, perché caratterizzano la VC
  - Come valori da aspettarsi (media, mediana, moda)
  - Come dispersione
- Parametri di distribuzioni notevoli
  - Distribuzioni note e studiate
  - Collegate a fenomeni naturali o comuni
  - O pratiche come strumenti di stima
- Esempi
  - Esponenziale, parametro: $\lambda = 1/\mu$
  - Normale, parametri: μ (moda, mediana) e σ
  - Triangolare, parametri: min, max e moda

Riferimenti

- G. Gallo, *Note di Simulazione*, cap. 3.1
- M. Pidd, Computer Simulation in Management Science, Cap. 11, J. Wiley & Sons, 1998